Thứ năm, 25/07/2024 15:07

Các nhà khoa học xã hội nên định hướng bản thân như thế nào trong kỷ nguyên trí tuệ nhân tạo?

Công nghệ trí tuệ nhân tạo (AI) đang mở ra nhiều triển vọng trong nghiên cứu khoa học, nhưng cũng đặt ra những thách thức về tính chính xác, sự thiên lệch thông tin và vấn đề đồng thuận trong sử dụng dữ liệu. Thay vì né tránh hoặc ngăn cản sự phát triển này, các nhà nghiên cứu nên chủ động thích ứng và tích hợp AI vào quy trình làm việc một cách có trách nhiệm, đồng thời duy trì tư duy phản biện và nền tảng lý thuyết vững chắc. Một hướng tiếp cận đầy tiềm năng là kết hợp lý thuyết sâu như Mindsponge với phương pháp tính toán như suy luận Bayes. Cách tiếp cận này có thể là chìa khóa giúp các nhà khoa học xã hội tiến hành nghiên cứu sáng tạo, hiệu quả và có trách nhiệm, bước vào một kỷ nguyên AI do con người dẫn dắt, thay vì bị cuốn vào một môi trường hỗn loạn do AI tạo ra.

Một trong những môn học khiến nhiều học viên cảm thấy thử thách nhất chính là môn Tư tưởng xã hội, vì họ phải đọc và viết tóm tắt cho ít nhất 5 bài báo khoa học hoặc một cuốn sách (thường dày từ 100-250 trang) vào 2 buổi trong tuần. Trên lớp, giáo sư thường yêu cầu mỗi học viên lần lượt đưa ra quan điểm cá nhân về bài đọc, tự đặt câu hỏi và cùng thảo luận về các vấn đề xã hội liên quan. Sau mỗi buổi học, học viên phải nộp một bài luận trình bày góc nhìn của bản thân về nội dung bài đọc và cuộc thảo luận trên lớp.

Ngày nay, mọi thứ đã thay đổi khi bài tập về nhà của các học viên đối với môn Tư tưởng xã hội hay bất kỳ môn học nào đã đều trở nên dễ dàng hơn rất nhiều nhờ vào AI. AI có thể tóm tắt bất cứ cuốn sách hay bài báo nào chỉ trong vài giây, rút ngắn thời gian làm việc [1]. Hơn thế nữa, AI còn có khả năng tổng hợp văn bản để tìm ra khoảng trống trong tri thức, tạo câu hỏi nghiên cứu, viết tổng quan nghiên cứu, phân tích dữ liệu định tính, thiết kế bài thuyết trình và thậm chí tự tạo nội dung bài đăng truyền thông [2]. Không quá khi nói rằng, AI đang dần trở thành một “thần đèn” trong thế giới học thuật. Riêng trong lĩnh vực tâm lý học, xã hội học và truyền thông, tính đến năm 2023, có khoảng 250 ứng dụng AI đã được tạo ra để hỗ trợ các nhà nghiên cứu làm khoa học [3]. Các mô hình ngôn ngữ lớn đã được sử dụng để hỗ trợ viết giả thuyết khoa học, tổng hợp thông tin đánh giá nghiên cứu và thực hiện phân tích dự đoán. Trong một cuộc khảo sát gần đây được công bố bởi *Turnitin*, gần 800 trong số 1600 sinh viên sau đại học và 200 trong số 1000 giảng viên cho biết, họ thường xuyên sử dụng AI cho nghiên cứu và học tập.

Tuy nhiên, bên cạnh những lợi ích to lớn, AI cũng tiềm ẩn một số rủi ro và hạn chế đáng quan ngại mà các nhà khoa học cần thận trọng xem xét. Nổi bật trong số đó là hiện tượng ảo giác, sự thiên lệch góc nhìn đối với thông tin và những thách thức liên quan đến sự đồng thuận trong việc sử dụng dữ liệu [4]. Khi lạm dụng hoặc phụ thuộc quá mức vào AI, 3 rủi ro kể trên có thể ảnh hưởng nghiêm trọng đến tính chính xác, độ tin cậy của kết quả và đạo đức khoa học [5].

Hiện tượng ảo giác là một trong những thách thức đáng chú ý nhất khi sử dụng AI, đặc biệt là các mô hình ngôn ngữ lớn. Đây là tình huống khi AI tạo ra thông tin sai lệch, bịa đặt dữ liệu, hoặc đưa ra những phản hồi

không có cơ sở thực tế. Một ví dụ điển hình xảy ra vào năm ngoái, khi ChatGPT đã tạo ra một tin giả về việc một giáo sư đại học có thật ngoài đời bị cáo buộc quấy rối tình dục, thậm chí còn bịa đặt cả trích dẫn từ tờ *Washington Post*. Để hiểu rõ nguyên nhân của hiện tượng này, chúng ta cần đi sâu vào cơ chế hoạt động của các mô hình ngôn ngữ lớn.

Mô hình ngôn ngữ lớn là một phần của mô hình nền tảng, được huấn luyện trên một lượng lớn dữ liệu khổng lồ không gán nhãn và tự giám sát. Quá trình này cho phép mô hình học hỏi từ các mẫu trong dữ liệu, tạo ra kết quả có khả năng tổng quát hóa và thích ứng cao. ChatGPT là một ví dụ tiêu biểu của mô hình ngôn ngữ lớn, có thể tạo ra văn bản gần giống con người. Thay vì tra cứu thông tin từ một cơ sở dữ liệu cụ thể, ChatGPT hoạt động bằng cách dự đoán từ tiếp theo trong câu, dựa trên mối quan hệ xác suất thống kê giữa các từ xung quanh. Kết quả là, mặc dù ChatGPT có thể tạo ra câu trả lời đúng về mặt cú pháp, nhưng không nhất thiết chính xác về mặt ngữ nghĩa hoặc ngữ cảnh. Giải thích đơn giản là ChatGPT không thực sự "hiểu" ý nghĩa sâu xa của câu hỏi hoặc ngữ cảnh của cuộc đối thoại, mà chỉ dựa vào các mẫu thống kê đã học được trước đó. Hiện tưởng ảo giác trong AI đang trở thành một vấn đề ngày càng nghiêm trọng, đặc biệt khi con người có xu hướng tin tưởng vào kết quả từ máy móc hơn là từ nguồn thông tin của con người.

Rủi ro thứ hai là sự thiên lệch góc nhìn đối với thông tin. Một nghiên cứu đã chỉ ra hạn chế này khi các tác giả sử dụng các mô hình ngôn ngữ lớn để phân tích báo cáo tài chính [6]. Kết quả cho thấy, độ chính xác của GPT-3 trong việc đưa ra lý luận đúng chỉ đạt dưới 50%. Hiệu suất kém này có thể do GPT-3 chưa từng tiếp xúc với mô hình tương tự như yêu cầu họ đưa ra trong quá trình huấn luyện. Trong lĩnh vực khoa học xã hội, sự thiên lệch này có thể dẫn đến những kết quả và phân tích không công bằng và thiếu khách quan. Các nhà khoa học cảnh báo rằng, khi các nhà nghiên cứu không nhận thức đầy đủ về sự thiên lệch này, họ có thể rơi vào “ảo tưởng về tính khách quan" [4]. Nói cách khác, họ có thể lầm tưởng rằng các công cụ AI không mang quan điểm riêng hoặc có khả năng đại diện cho mọi góc nhìn. Trong khi thực tế, các công cụ này đều chứa đựng những quan điểm bắt nguồn từ dữ liệu huấn luyện và từ chính những nhà phát triển chúng.

Vấn đề cuối cùng nhưng cũng không kém phần quan trọng là sự đồng thuận trong việc sử dụng dữ liệu. Điều này xảy ra do nhiều công cụ AI được phát triển bởi các công ty tư nhân với dữ liệu huấn luyện và mô hình đóng, khiến người dùng không thể xác minh nguồn gốc và tính hợp pháp của dữ liệu được sử dụng [5]. Câu hỏi đặt ra là liệu những dữ liệu này có được thu thập với sự đồng ý của chủ sở hữu và tuân thủ các quy định về bản quyền hay không? Đây là một trong những mối quan ngại hàng đầu của người dùng, bởi các phản hồi họ nhận được từ các mô hình ngôn ngữ lớn có thể được trích xuất từ những nguồn nhạy cảm như tin nhắn riêng tư hoặc thậm chí là thông tin từ trẻ vị thành niên mà không có sự cho phép. Để đảm bảo tính minh bạch và đạo đức trong nghiên cứu, các nhà khoa học cần phải chủ động tìm hiểu và giải quyết những thách thức này khi sử dụng AI như một công cụ nghiên cứu.

Ngoài những rủi ro và mối lo ngại đã đề cập, việc phụ thuộc quá mức vào công cụ AI còn tiềm ẩn những tác động tiêu cực khác đối với cộng đồng học thuật. Một trong những hệ lụy đáng quan ngại nhất là sự suy giảm khả năng tư duy phản biện của sinh viên và nguy cơ góp phần vào cuộc khủng hoảng tái lập trong nghiên cứu khoa học. Một nghiên cứu trên sinh viên đại học ở Pakistan và Trung Quốc cho thấy, AI có tác động đáng kể đến việc suy giảm khả năng ra quyết định (27,7% sinh viên) và gây ra tình trạng lười biếng (68,9% sinh viên) [7]. Hơn nữa, trong bối cảnh AI đang phát triển nhanh chóng và khó dự đoán, việc sử dụng công nghệ AI như một công cụ phân tích có thể đặt các nhà nghiên cứu vào tình thế bấp bênh [5]. Do

các nhà khoa học có thể gặp khó khăn trong việc tái tạo lại chính nghiên cứu của mình, điều này làm trầm trọng thêm các thách thức hiện có trong giới học thuật về việc duy trì niềm tin của công chúng vào khoa học [8].

Trong bối cảnh AI đang ngày càng phát triển và ảnh hưởng sâu rộng đến lĩnh vực nghiên cứu, một câu hỏi quan trọng được đặt ra: Làm thế nào để chúng ta xây dựng mối quan hệ hiệu quả và có trách nhiệm với công cụ mạnh mẽ này? Đồng thời, các nhà khoa học nên định hướng bản thân như thế nào trong kỷ nguyên AI?

Thay vì có thái độ phòng thủ hoặc cố gắng ngăn chặn sự phát triển của AI, chúng tôi cho rằng, cộng đồng học thuật - bao gồm các nhà nghiên cứu, biên tập và người đánh giá nên chủ động thích ứng và tích hợp công nghệ AI vào quy trình làm việc của mình một cách thông minh, có chọn lọc, và có trách nhiệm. Lý do là vì AI đang không ngừng phát triển và tự hoàn thiện mỗi ngày. Một minh chứng rõ ràng cho sự tiến bộ này là phiên bản cập nhật ChatGPT 4, với khả năng đưa ra các phản hồi hợp lý và logic vượt trội so với phiên bản tiền nhiệm ChatGPT 3 trong việc giải quyết các bài toán phức tạp [9]. Những điểm mạnh và điểm yếu của AI giống như hai mặt của một đồng xu: sức mạnh của AI trong việc tìm kiếm, tổng hợp và tạo ra thông tin có thể trở thành lợi thế to lớn khi chúng ta biết cách khai thác hiệu quả; ngược lại, nếu không kiểm soát được hoặc phụ thuộc quá mức vào AI, những ưu điểm này có thể nhanh chóng biến thành điểm yếu, ảnh hưởng tiêu cực đến chất lượng nghiên cứu và sự phát triển của tư duy độc lập.

Để thích ứng và áp dụng AI một cách hiệu quả, các nhà nghiên cứu, biên tập viên và người đánh giá cần có nền tảng lý thuyết sâu sắc cho tư duy khái niệm và khả năng xử lý, kết hợp và xác minh thông tin do AI cung cấp. Điều này đòi hỏi việc tích hợp lý thuyết sâu và các khung phân tích để xây dựng nền tảng logic và lý luận vững chắc trong mọi nghiên cứu.



Cuốn sách phương pháp luận Bayesian Mindsponge Framework.

Trong khoa học xã hội và nhân văn, việc tích hợp lý thuyết sâu như Mindsponge với phương pháp tính toán như suy luận Bayes mở ra một hướng đi đầy hứa hẹn. Lý thuyết Mindsponge tập trung vào cấp độ cơ bản

của tâm lý và hành vi con người thông qua lăng kính xử lý thông tin, giúp hiểu sâu sắc hơn về cách con người tương tác với thông tin và môi trường xung quanh. Hơn nữa, lý thuyết Mindsponge không mâu thuẫn với các lý thuyết và khung tâm lý và xã hội hiện có, mà thay vào đó, nó mở rộng, giải quyết các mâu thuẫn và kết nối các khái niệm thông qua góc nhìn động về xử lý thông tin [10].

Khi kết hợp Lý thuyết Mindsponge với tư duy tính toán như suy luận Bayes, các nhà nghiên cứu có thể tận dụng cả sức mạnh lý luận lý thuyết và ưu điểm thống kê để nâng cao hiệu quả và độ chính xác của nghiên cứu [11, 12]. Ví dụ, Lý thuyết Mindsponge giúp thiết kế phương pháp khảo sát hiệu quả hơn bằng cách xác định chính xác các yếu tố liên quan đến quá trình tâm lý trong nghiên cứu, từ đó cho phép áp dụng các phép đo và loại dữ liệu phù hợp [11]. Ngoài ra, suy luận Bayes bổ sung thêm nhiều lợi thế, như khả năng đưa thông tin đáng tin cậy vào trước khi phân tích dữ liệu, giải quyết vấn đề đa cộng tuyến trong phân tích thống kê. So với phương pháp tần suất truyền thống, suy luận Bayes còn giúp các nhà khoa học dễ dàng xác định vùng tin cậy chứa giá trị tham số thực với xác suất cao [10]. Sự kết hợp này tạo nên phương pháp luận Bayesian Mindsponge Framework (BMF), một công cụ đa năng đã được ứng dụng rộng rãi trong nhiều lĩnh vực, từ tâm lý học, giáo dục đến tâm lý môi trường và chăm sóc sức khỏe. BMF không chỉ nâng cao chất lượng nghiên cứu mà còn mở ra những hướng tiếp cận mới, sáng tạo trong việc giải quyết các vấn đề phức tạp trong khoa học xã hội và nhân văn [11].

Vì AI không "suy nghĩ" như con người và vẫn có những hạn chế trong cách thức hoạt động, cộng đồng học thuật, bao gồm các nhà nghiên cứu, biên tập và đánh giá cần có nền tảng lý thuyết sâu sắc để tư duy phản biện và sáng tạo. Dựa trên những lập luận trên, lý thuyết Mindsponge và phương pháp luận BMF có tiềm năng mang lại một con đường hứa hẹn cho các nhà khoa học xã hội và nhân văn để tiến hành nghiên cứu một cách sáng tạo, có trách nhiệm và hiệu quả. Bằng cách tận dụng các ưu điểm của AI trong khi duy trì khả năng kiểm soát và đánh giá của con người dựa trên nền tảng lý thuyết vững chắc, chúng ta có thể bước vào một kỷ nguyên AI do con người dẫn dắt, thay vì bị cuốn vào một môi trường hỗn loạn do AI gây ra.

TÀI LIỆU THAM KHẢO

[1] T.B. Brown, B. Mann, N. Ryder, et al. (2020), “Language models are few-shot learners”, *Arxiv*, DOI: 10.48550/arXiv.2005.14165.

[2] T. Davidson (2024), “Start generating: Harnessing generative artificial intelligence for sociological research”, *SocArXiv Papers*, DOI: 10.31235/osf.io/u9nft.

[3] M.S. Richardson, L. Brown, M. Paul, et al. (2023), “Artificial intelligence applications for social science research”, *Scholars Junction*, Mississippi State University.

[4] L. Messeri, M.J. Crockett (2024), “Artificial intelligence and illusions of understanding in scientific research”, *Nature*, **627**, pp.49-58, DOI: 10.1038/s41586-024-07146-0.

[5] A. Spirling (2023), “Why open-source generative AI models are an ethical way forward for science”, *Nature*, **616**, pp.413-413, DOI: 10.1038/d41586-023-01295-4.

[6] Z. Chen, S. Li, C. Smiley, et al. (2022), CONVFINQA: Exploring the chain of numerical reasoning in conversational finance question answering”, *Proceedings of The 2022 Conference on Empirical Methods in Natural Language Processing*, pp.6279-6292.

[7] S.F. Ahmad, H. Han, M.M. Alam, et al. (2023), "Impact of artificial intelligence on human loss in decision making, laziness and safety in education", *Humanit. Soc. Sci. Commun.*, **10(1)**, DOI: 10.1057/s41599-023-01787-8.

[8] M.T. Ho, Q.H. Vuong (2019), "The values and challenges of "openness" in addressing the reproducibility crisis and regaining public trust in social sciences and humanities", *European Science Editing*, **45(1)**, pp.14-17, DOI: 10.20316/ESE.2019.45.17021.

[9] Q.H. Vuong, V.P. La, M.H. Nguyen, et al. (2023), "Are we at the start of the artificial intelligence era in academic publishing?", *Science Editing*, **10(2)**, pp.158-164, DOI: 10.6087/kcse.310.

[10] Q.H. Vuong (2023), *Mindsponge Theory*, Sciendo, 256pp.

[11] Q.H. Vuong, M.H. Nguyen, V.P. La (2022), *The Mindsponge and BMF Analytics for Innovative Thinking in Social Sciences and Humanities*, Sciendo, 430pp.

[12] M.H. Nguyen, V.P. La, T.T. Le, et al. (2022), "Introduction to Bayesian Mindsponge Framework analytics: An innovative method for social and psychological research", *MethodsX*, **9**, DOI: 10.1016/j.mex.2022.101808.

Nguồn:

[1]*Đại học Công nghệ Michigan, Mỹ*

[2]*Bộ Khoa học và Công nghệ*

[In bài viết] [Đóng]